用 CLEAN 算法对嫦娥四号低频射电频谱仪信号进行干扰抑制

刘晨迪[1,2,3]，周建锋[4]，苏彦[1,2,3]

（1.中国科学院国家天文台，北京 100101；2.中国科学院大学，北京 100049；3.中国科学院月球与深空探测重点实验室，北京 100101；4.清华大学工程物理系，北京 100084）

摘 要：嫦娥四号低频射电频谱仪放置在月球背面，其天然观测条件得天独厚。然而，嫦娥四号平台存在约 10^{-15}W/(m^2*Hz)级别的强干扰，并且干扰在每道时域数据中都存在明显的差异，这大大削弱了低频射电频谱仪的天文观测灵敏度。为此，本论文从两组信号的相关性出发，提出基于 CLEAN 算法，借助互相关功率谱、傅立叶级数展开等工具，把低频射电频谱仪 A、B、C 天线的时域观测数据切分为强相关的 CLEAN 模型信号和部分相关的残余信号两个部分。其中 CLEAN 模型信号主要由平台干扰信号和可能的低频强射电爆发组成；残余信号由接收机噪声、未扣除干净的平台干扰信号和常规的低频射电信号组成。将该方法应用到实际数据中，结果表明嫦娥四号低频射电频谱仪的未积分灵敏度可以提高约 8 个数量级，达到 10^{-23}W/(m^2*Hz)水平。在此基础上，基于对平台干扰信号中确定性成分和宽带随机成分的分别处理，再借助低频射电爆发信号和平台干扰信号在功率谱上不同的表现，以及常规低频射电天文信号受月球自转调制等信息，将来科学分析工作的重点是进一步处理 CLEAN 模型信号和残余信号，以期发现低频强射电天文爆发信号，乃至对全天区进行粗略的成像。

关键词：嫦娥四号低频射电频谱仪，平台干扰分离，CLEAN 算法。

中图分类号：P16　　　　文献标识码：A　　　　文章编号：

1　引言

嫦娥四号低频射电频谱仪（Low frequency radio spectrum analyzer 简称 LFRS）放置在月球背面。由于月球屏蔽了大量来自地球的无线电干扰，因此 LFRS 具有得天独厚的观测条件，可以对太阳等宇宙天体的射电信号进行观测。[1]但是，由于着陆器平台干扰信号比接收到的低频射电天文信号大很多，如果不分离平台干扰信号，将无法得到有效的低频射电天文信号数据。因此，需要提出一种有效分离平台干扰信号的方法，从而实现对干扰的抑制。

在射电天文领域，传统抑制干扰的方法有很多，包括对空白脉冲干扰的时间限制[2]，自适应频率选择滤波[3]，参数信号相减[4]，自适应波束成形[5]，时间自适应滤波[6]，带有参考天线的频域后相关处理[7]，子空间投影的空间滤波[8]等等。到目前为止，运用比较广泛的算法是单通道总功率变化检测器，它使得相关器的输出得到减小。Friedman（1996）基于对功率变化的检测，在 RATAN600 上实现了改进的功率检测器[9]。Weber 等人（1997 年）提出在所有滞后的情形下使用量化相关性来测试干扰的存在[10]。Maslakovic 等人 1996 年提出了一种基于小波分解的检测器[11]。这些都是单通道检测器，它们不利用干扰的空间特性分析。Kasper，Chute 和 Routledge 在 1982 年提出了将多个望远镜结合使用以改善探测质量的方法，此方法用于低频干涉测量，在这之中还提出了一种基于交叉频谱时间行为的鲁棒数据检查方法[12]。Barnbaum＆Bradley（1998）考虑了自适应过滤技术，他们提议使用参考天线和最小均方（LMS）类型算法消除 Green-Bank 射电望远镜的干扰[13]。W. A. BAAN 提出了一种针对韦斯特博克综合射电望远镜的射电干扰消除子系统[14]。这些都是在射电干扰消除方面典型的例子。

上面提到的射电干扰信号消除算法主要分为三类：第一类试图去通过线性的方式来区分射电干扰信号的特点，比如奇异值分解（SVD）[15]，主成分分析（PCA）[16]等等。如果射电干扰在时间和频率上表现出重复的模式，这些方法效果很好，但无法处理更多随机信号，例如由卫星引起的不规则信号。第二类是基于阈值的算法，例如 CUMSUM[8]和 SUMTHRESHOLD[4]，尽管它们很简单，但是这些方法相当可靠并且非常有效。第三类使用传统的监督机器学习技术（例如 K 最近邻和高斯混合模型）对射电干扰信号进行聚类[17]。

嫦娥四号 LFRS 所收到的平台干扰信号很强，和太阳爆发峰值的量级 10^{-15}W/(m^2*Hz)相当，

比普通的太阳爆发信号 10^{-18}W/(m^2*Hz)大三个量级，比太阳宁静时的信号 10^{-22}W/(m^2*Hz)（1S.F.U）大七个量级。之前仪器设计时推荐使用的谱减法[18]，虽然可以将信号的信噪比提升约 30dB，但是仍然不足以探测有效的射电天文信号。为此，本文提出了基于 CLEAN 算法的平台干扰信号分离方法，利用 LFRS 三根天线同时观测信号存在相关性的特点，并结合天线之间数据的互相关功率谱、傅立叶级数展开等工具，有效地分离平台干扰信号，从而实现对干扰的抑制。

本文的第 2 小节主要介绍 LFRS 的工作原理及接收信号构成成分，第 3 小节介绍 LFRS 平台干扰信号分离的方法及算法流程，第 4 小节展示新算法对 LFRS 实际数据的处理结果及效果分析，第 5 小节是讨论，最后的第 6 小节是结论。

2　LFRS 的工作原理及接收信号的构成

2.1 LFRS 简要介绍

太阳和宇宙中的其它来波信号是一个矢量信号。根据矢量场的理论，来波信号可以分解为任意三个互相垂直方向的分量，只要将这三个分量接收下来，进行合成，就可以得到矢量的大小和方向。通过后期处理，可以获得电场的频谱和时变信息。

LFRS 的天线分布如图 1 所示：三分量有源天线接收宇宙的来波电场。三根天线分别标注为“天线 A”、“天线 B”和“天线 C”，两两正交，还有一根短的“天线 D”，主要用于接收平台干扰信号。为天线数据处理做一个参考。

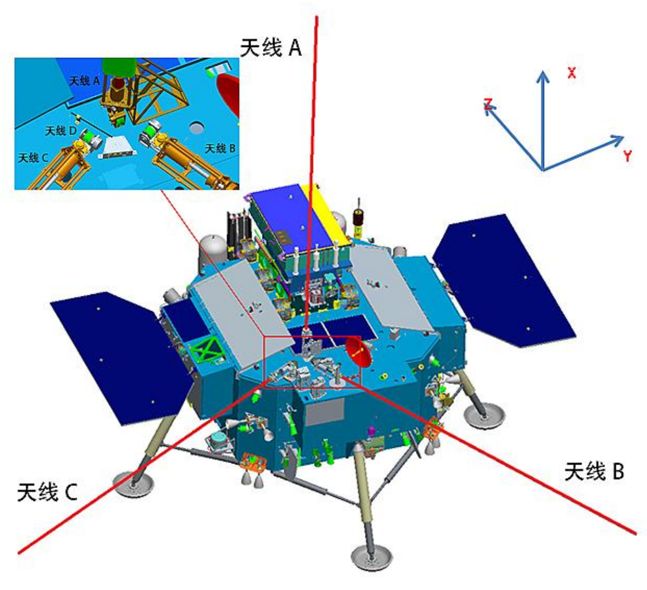


图 1 着陆器低频射电频谱仪的天线位置分布

Fig.1 Antenna location distribution of lander low-frequency radio spectrum analyzer

LFRS 具有高频和低频两个频率接收波段，高频对应的频率范围是 1-40MHz，低频对应的频率范围是 0.1-2MHz。接收机灵敏度为 $6nV/\sqrt{Hz}$，高频的频率分辨率为 100kHz，低频的频率分辨率为 5kHz。

LFRS 主要由四根接收天线、前置放大器、电子学单元、电缆组件等部分组成。有源天线安装于卫星本体外，其余组件安装于卫星本体内。电子学单元由控制器、配电器、基准时钟模块、四通道接收机、内定标模块、存储单元、通讯接口、数传接口等组成。LFRS 内部各单元之间及与数管分系统、供配电分系统之间的接口关系如图 2 所示。

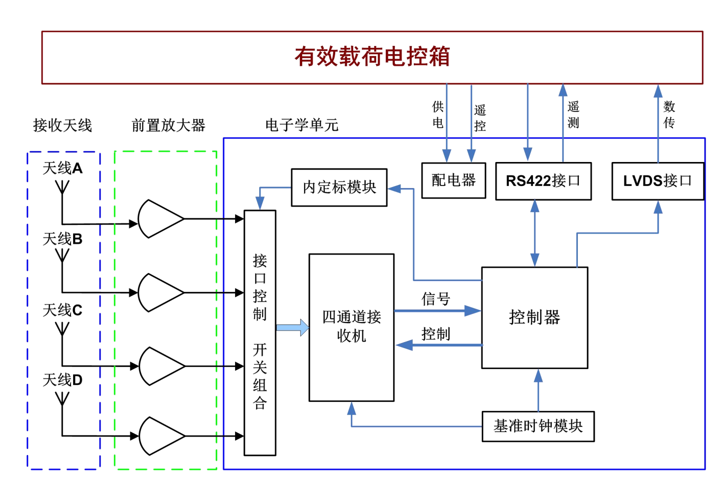


图 2 低频射电频谱仪系统组成及接口关系示意图

Fig. 2 Schematic diagram of low-frequency radio spectrum analyzer, with system composition and interface relationship

LFRS 有六种工作模式，分别为 1）待机工作模式 2）内定标工作模式 3）时频对比工作模式 4）频谱巡查工作模式 5）三天线工作模式 6）时域工作模式。其中运用最广泛的是三天线工作模式，在此模式下，LFRS 分别采集四根天线四个通道的科学数据，本文进行读取处理的数据均对应三天线工作模式。

LFRS 的科学数据分为 0A 级、0B 级、01 级、2A 级、2B 级、2C 级共六级。其中预处理数据经过分路解帧和分包，生成 0A 级数据；0A 级数据经过排序、去重复、两站优化拼接、去源包包头、根据采样周期形成 LFRS 数据块，生成 0B 级数据产品；多个 0B 级数据文件拼合成一个探测周期，生成 01 级数据产品；对 01 级数据进行傅立叶变化、定标等操作，生成 2A 级数据产品；对 2A 级数据加入太阳位置等信息，生成 2B 级数据产品；对 2B 级数据进行对消噪声处理，生成 2C 级数据产品。其中 1 级数据为时域数据，2 级数据为频域数据。本论文处理所用的原始数据均为 1 级数据。

2.2 接收信号的构成成分

LFRS 收到的信号主要由三部分构成，一部分是平台干扰信号，一部分是接收机噪声，最后是低频射电天文信号。

平台干扰信号的特点是幅度相对较强，两两天线之间的相关性高。如果干扰源位置确定，各天线接收信号之间的强度比例不随时间变化。且来自电子设备的干扰信号，一般可分为固定频率的发射信号，如通信发射机，时钟信号等，另一部分是宽带的随机噪声信号，这是由于 EMC 兼容问题带来的。最终的平台干扰信号如图 6 所示，既有确定性变化的成分，也有宽带随机性噪声成分。

低频射电天文信号也具有一定的相关性，除此之外，还有一些天体源所特有的性质。比如，低频太阳射电信号其具有瞬时性和频谱漂移等特点，为此其具有很强的频谱可识别特征[19]，如图 3 所示。再比如，快速射电爆（Fast Radio Burst, FRB）FRB200428 的持续时间约 0.61ms，能量通量（Fluence）达到 1.5 MJy ms[20]，其辐射强度足以被 LFRS 探测到。另外，由于仪器指向随月球的自转而改变，低频射电天文信号在 LFRS 各天线之间的强度比例随时间呈现有规律的变化。所有这些特性，都有助在强平台干扰的背景下，识别有用的低频射电天文信号。

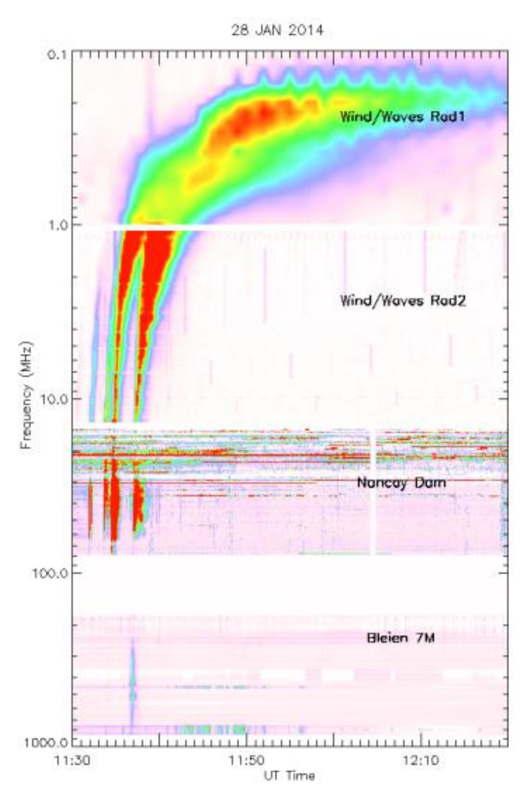


图 3:低频太阳射电信号频谱

Fig.3 Low frequency spectrum of solar radio signal

接收机噪声的特点是天线两两之间无相关性。可以利用该特性，通过计算天线之间数据的相关性来评估平台干扰信号的分离程度。

图 4 展示了 LFRSB 天线观测到的一个频谱实例。由图可以得知，频谱在 1MHz 频率附近的流量密度大约为10^{-15}W/(m^2*Hz)，与太阳爆发峰值流量密度10^{-15}W/(m^2*Hz)相当，比太阳爆发流量密度10^{-18}W/(m^2*Hz)大了 2-3 个量级，说明着陆器平台干扰比我们希望探测的太阳爆发信号要大许多。如果不通过一定的方法分离平台干扰信号，将无法有效地提取出想要的低频射电天文信号。

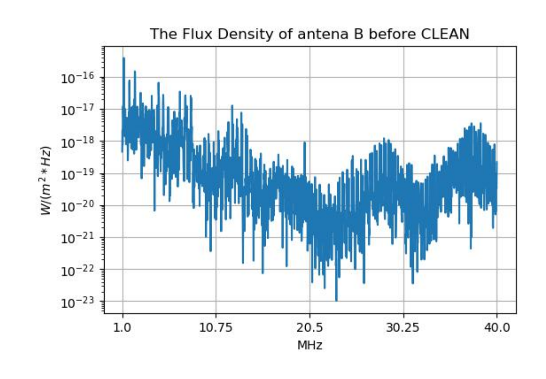


图 4 B 天线采集到的高频波段的一个频谱实例，对应第 23 月昼第二道数据

Fig.4 An example of the spectrum of antenna B at the high-frequency band，corresponding to the second channel data on the 23rd moon day.

3 LFRS 平台干扰信号的分离方法

低频射电源与平台干扰源会在 LFRS 的 3 个天线上激发出互相关联的接收信号。本论文的目的是从两组信号的相关性出发，借助 CLEAN 方法，以及互相关功率谱、傅立叶级数展开等工具，把 LFRS 观测信号中的强相关成分与部分相关成分进行分离，以便进行后续的科学

分析。分离之后，原始的 LFRS 时域信号被切分为两个成分：即强相关的 CLEAN 模型信号和部分相关的残余信号。在这个过程中，原始信号的信息没有丢失或损伤。

CLEAN 模型信号主要由平台干扰信号组成。当低频强射电爆发信号到来时，也会进入 CLEAN 模型信号。由于数传的限制，LFRS 大约以 1 秒钟为间隔采集一道时域数据，采集间隔并不严格相等。每道数据由 4096 个数据组成，采样频率为 100MHz，采样持续时间只有大约 40 微秒。因此，不同道时域数据有明显的差别（参见图 6 与图 7）。

为了寻找有效的低频强射电爆发信号，需要提取每一道时域数据的 CLEAN 模型信号，并对其中的确定性成分和宽带随机成分分别进行整理。例如，对于 CLEAN 模型信号中的确定性时域变化结构或其频谱成分，可以把它们当作某种系统误差，同样借助 CLEAN 算法，将其扣除。剩下的 CLEAN 模型信号，由宽带随机信号组成，可以通过平均多道数据的方式，降低其涨落水平，提高探测低频强射电天文爆发的灵敏度。

残余信号主要由不相关的接收机噪声、未分离出去的平台干扰信号、以及常规低频射电天文信号组成。残余信号的特征类似于随机白噪声，里面没有确定性的信号成分，可以借助于数理统计工具来进行分析。例如，可以借鉴微波辐射计的思想，通过积分平均的方式进一步提高残余信号的探测灵敏度。

我们的最终目标是用嫦娥四号 LFRS 的观测数据来搜索低频射电天文信号，并进行深入的科学研究。此时，无论是 CLEAN 模型信号还是残余信号都是缺一不可的。

时标短（分钟到天量级）、幅度强（MJy 量级）的低频强射电爆发往往隐藏在 CLEAN 模型信号中。对这类射电爆发信号的提取离不开宁静状态的 CLEAN 模型信号的分类整理。特别强的低频射电天文爆发，在单道数据中就有明显异常的表现。此时，可以借助交叉相关，直接扣除该道数据对应的宁静状态的 CLEAN 模型信号，获得射电爆发信号的时域、频谱特征。如果射电爆发强度与平台干扰中的宽带随机成分接近，那么先扣除 CLEAN 模型信号中的确定性成分，再通过多道数据积分平均的方式，提高探测灵敏度，最终获得这类低频射电天文爆发信号的时域、频谱特征。

一些时标长（超过月量级）、幅度弱的低频射电天文信号则隐藏在残余信号中。为了探测这类弱信号，首先要借助星载定标装置或其它外部稳定的定标源，确认接收机工作稳定。之后，可以直接从残余信号出发，同样借助微波辐射计的思路，通过积分平均，提高残余信号的探测灵敏度。理论上讲，LFRS 三个天线残余信号的积分值会随着月球自转而发生调制变化。这样的调制信息甚至可以用来对全天进行粗略的成像。

3.1 用 CLEAN 算法分离平台干扰信号的基本思路

设 LFRS 中一个天线收到的信号为 $S(t)$，它包含了平台干扰信号 $I(t)$，宇宙射电信号 $C(t)$，以及接收机噪声 $N(t)$。这些信号之间的关系如下：

$$\begin{aligned} S_A(t) &= \alpha_A(t)I(t) + \beta_A(t)C(t) + N_A(t) \\ S_B(t) &= \alpha_B(t)I(t) + \beta_B(t)C(t) + N_B(t) \\ S_C(t) &= \alpha_C(t)I(t) + \beta_C(t)C(t) + N_C(t) \end{aligned} \quad (1)$$

其中$S_A(t)$、$S_B(t)$、$S_C(t)$分别为天线 A、B、C 接收到的信号；$N_A(t)$、$N_B(t)$、$N_C(t)$分别为天线 A、B、C 对应接收机的本地噪声，它们之间相互独立；$\alpha_A(t)$、$\alpha_B(t)$、$\alpha_C(t)$为平台干扰信号在天线 A、B、C 上的投射系数，$\beta_A(t)$、$\beta_B(t)$、$\beta_C(t)$为宇宙射电信号在天线 A、B、C 上的投射系数。

在 2.2 小节中提到，平台干扰信号比低频射电天文信号大约 3～10 个量级。因此，分离平台干扰是进行有效天文观测的一个基本前提。

经典的 CLEAN 算法在脏图或残余图搜索最大值，此位置处包含部分真实目标，用减点源扩散函数（PSF）的方式将这部分真实目标提取出来，如此不断迭代，直到残余图中只剩噪声结构为止[21]。CLEAN 算法同样可以应用到时域信号上，比如利用 CLEAN 方法对时间序列进

行频谱分析[22]，利用时域 CLEAN 识别声源[23]，以及利用 CLEAN 方法提取一些特定结构的信号，比如周期信号、准周期信号等等[24]。

我们利用 CLEAN 算法提取 LFRS 的平台干扰信号。算法的输入数据为一组（2 个以上）天线的观测数据。首先，计算每个天线的频谱，以及每一对天线的互相关功率谱，并将所有的互相关功率谱叠加起来。由于每个天线的接收机噪声 $N(t)$是相互独立的，因此互相关功率谱中的噪声水平会降低，而叠加后的互相关功率谱的信噪比会进一步增强，这有利于提取平台干扰信号。

叠加的互相关功率谱包含了平台干扰信号对应的频谱成分，同时囊括了振幅和相位信息。根据这些信息，我们借助 CLEAN 算法和傅立叶级数展开公式，可以逐一将某个频率成分对应的三角函数信号从时域数据中扣除。考虑到每个天线及其接收机的增益并不一致，在扣除时每个天线的 CLEAN 因子是不一样的。

离散傅立叶变化所对应的时域及频域信号都是周期信号。如果信号首尾的值并不一致，就会在傅立叶变换时出现频谱泄漏，导致频率估计出现偏差。通常，实际采集到的信号并不满足首尾连续的条件，正如图 6 中 LFRS 的 A、B、C 天线信号所示的那样。一般可以采用加窗函数的方式，来避免频谱泄漏。但是加窗函数后会导致信号发生扭曲，幅度值发生变化，假如我们直接采用傅立叶反变换的方式来进行信号分离，则会导致较大的偏差。为此，我们引入 CLEAN 方法，通过加窗函数的方式，来获得准确的频谱估计；然后在信号分离时，逐步扣除某一个谐波分量，可以避免扭曲信号，从而把首尾不连续的干扰信号精确的分离出来。

当大部分平台干扰信号被 CLEAN 算法提取出来后，残余信号主要由接收机噪声所主导，此时可以停止 CLEAN。来自宇宙的射电信号，其幅度往往小于噪声水平，加上小部分平台干扰信号，仍然隐藏在接收机噪声中。之后，可以通过对时域信号积分平均的手段，提高残余信号的信噪比，并利用信号随时间的变化特性来进一步辨别、分离平台干扰信号和低频射电天文信号，从而实现对干扰的抑制。

3.2 CLEAN 算法的流程

按照 3.1 子小节的思路，我们给出用 CLEAN 算法分离平台干扰信号的流程图。为了更简明清晰的展示算法，我们先考虑两根天线（分别记为 A 和 B）的情形，并且只考虑连续时域信号。定义在某一道观测数据中，天线 A 的时域数据记为$f^A(t)$，天线 B 的时域数据记为$f^B(t)$。

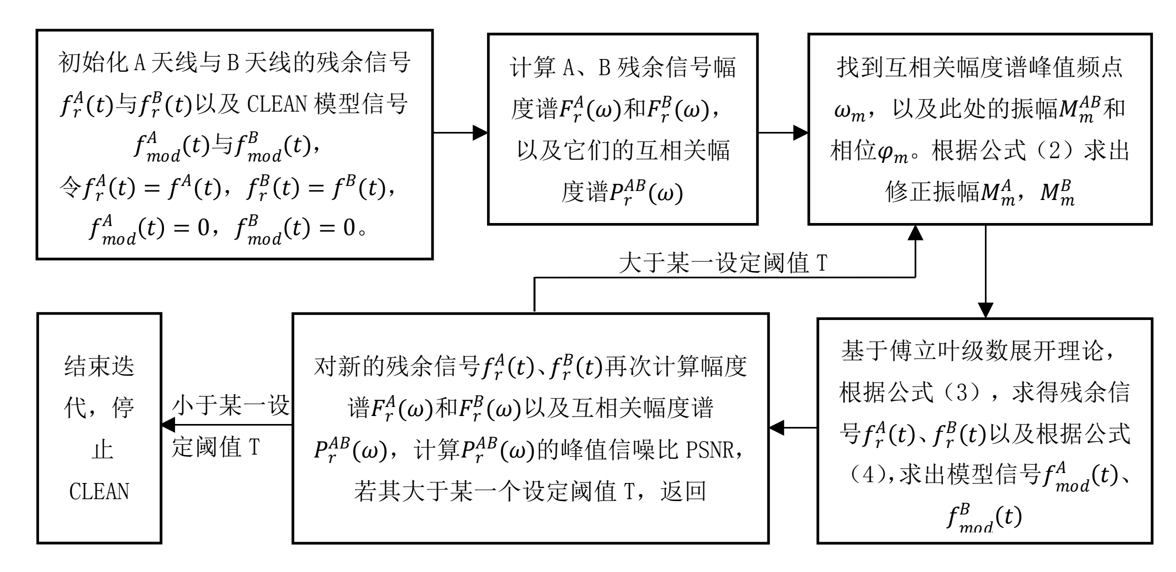


图 5 CLEAN 算法流程图

Fig.5 CLEAN algorithm flow chart

相关的公式如公式（2）、（3）、（4）所示

$$M_m^A=\frac{\overline{M^A}}{\overline{M^{AB}}}M_m^{AB},M_m^B=\frac{\overline{M^B}}{\overline{M^{AB}}}M_m^{AB} \tag{2}$$

其中$\overline{M^A}$和$\overline{M^B}$分别为$F_r^A(\omega)$与$F_r^B(\omega)$的平均振幅，$\overline{M^{AB}}$为$P_r^{AB}(\omega)$的平均振幅。

$$f_r^A(t)=f_r^A(t)-\delta M_m^A\cos(\omega_m t+\varphi_m)$$
$$f_r^B(t)=f_r^B(t)-\delta M_m^B\cos(\omega_m t+\varphi_m) \tag{3}$$
$$f_{mod}^A(t)=f_{mod}^A(t)+\delta M_m^A\cos(\omega_m t+\varphi_m)$$

$$f_{mod}^B(t)=f_{mod}^B(t)+\delta M_m^B\cos(\omega_m t+\varphi_m) \tag{4}$$

本算法使用了修正振幅M_m^A和M_m^B，没有直接采用频谱$F_r^A(\omega)$和$F_r^B(\omega)$在频点ω_m处的振幅，其原因是当信噪比较低时，频谱的振幅由噪声谱振幅主导，而互相关功率谱$P_r^{AB}(\omega)$在频点ω_m处的振幅M_m^{AB}受噪声的影响较小。

在这里，我们假定天线 A 和 B 的时域信号$f^A(t)$和$f^B(t)$是同步的。若由于线路补偿没有做好，两个时域信号之间存在时间延迟，那么需要通过交叉相关等方法计算该时间延迟，并对时域信号做延迟修正，然后再用 CLEAN 算法做平台干扰信号的提取。

通过 CLEAN 算法分离平台干扰信号之后，得到的残余信号主要成分是接收机噪声，以及隐藏其中的宇宙射电信号，还有少部分没有扣除干净的平台干扰信号。此时两个天线残余信号之间的相关系数比较低，例如绝对值小于 0.3。这样，通过计算信噪比发现，残余信号中的平台干扰信号和接收机噪声基本在同一个量级水平上。再往后，可以通过积分平均的方式来提高残余信号的信噪比，这与传统辐射计的工作思路一致。至于平台干扰信号和宇宙射电信号的区分，则要借助三个天线信号随频谱、时间和来波方向变化的规律。

4 LFRS 实例数据处理

4.1 LFRS 数据处理流程

LFRS 原始数据由嫦娥四号地面应用系统数据发布中心提供。数据放在嫦娥四号低频射电频谱仪载荷相关文件夹下，按照 0A、0B、01、2A、2B 进行分级排列，每一级数据均按照每一个月昼一个文件夹进行排列，每一个月昼通常有 14 个文件，其中 7 个是数据文件，7 个是数据采集参数文件。7 个数据文件对应的是在一个月昼中 7 次开机时间段所采集到的科学数据，我们在研究不同时间段的数据时，分别对这 7 个文件进行读取，并结合相应的 7 个数据采集参数文件，对每一个时段的采集参数进行确定。

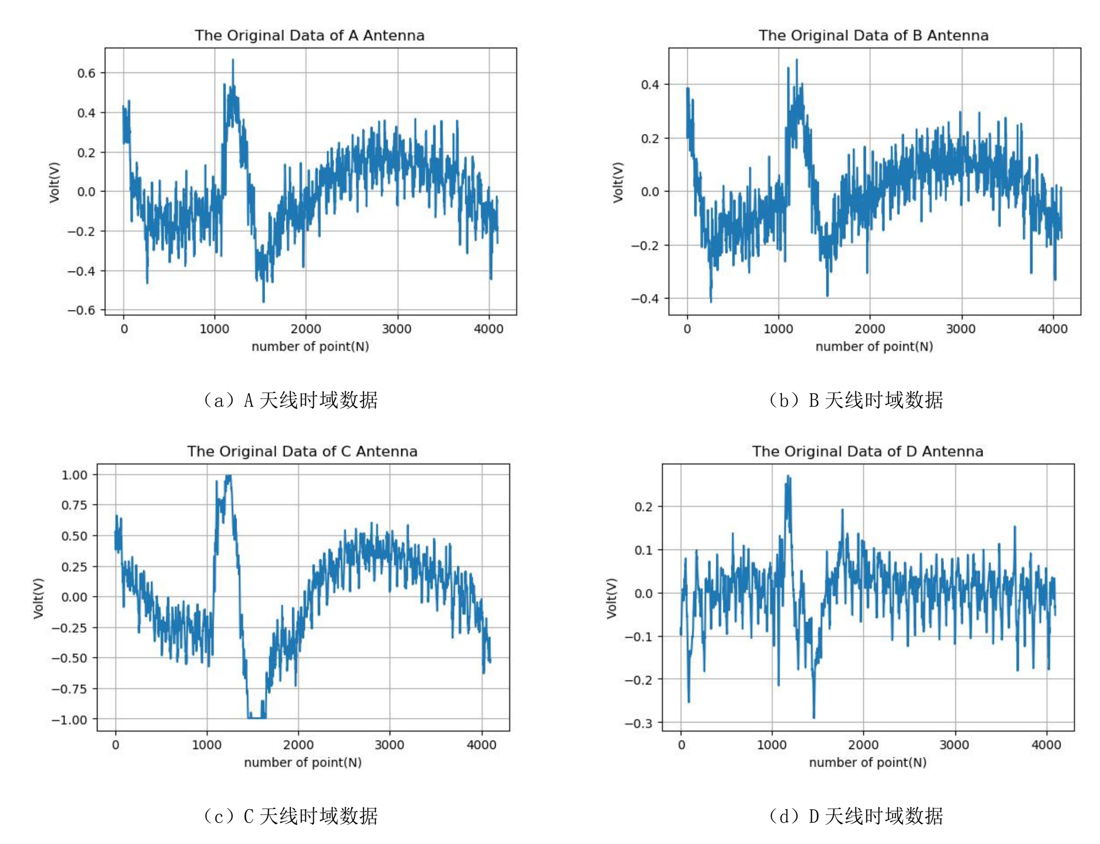
以第 23 月昼高频 1 级时域数据的第一道数据为例。对应的数据文件名为 CE4_GRAS_LFRS-TR_SCI_P_20201014020000_20201014140000_0135_B.01，其中 GRAS 代表地面应用系统，SCI 代表科学数据，0135 代表这个文件是总共第 135 个数据文件，数据结尾的 01 扩展名代表数据是 1 级数据，为时域数据。从文件名可以看出采集的数据时间段是从 2020 年 10 月 14 日 2 点 0 分 0 秒到 2020 年 10 月 14 日 14 点 0 分 0 秒，这个时间是世界时，转换为北京时间还需要加 8 小时进行换算。

在数据文件中，数据按照道的顺序依次排列，每一道数据由数据头（79 字节）和实际科学数据（8192 字节）组成。其中数据头包含了仪器的相关参数，科学数据共有 4096 个科学数据点，每个数据 2 个字节，由 16 位无符号整数组成，高字节在前，低字节在后。读取后还需要将整数电压转化为浮点数电压值，具体数据形式可参考文档《探月工程嫦娥四号任务着陆器低频射电频谱仪数据预处理方法设计》[25]。第一道数据一共采集了 4096 个点，高频数据的采样间隔为 0.01μs，低频数据的采样间隔为 0.2μs。

我们是通过 python 编程的方式，将数据按照道的顺序依次读取出来，将每一道 4096 个点排成一行，生成一个科学数据矩阵。处理的时候，提取每一行（每一道）数据，分别用第 3.2 子小节所述的 CLEAN 方法进行处理。

4.2 LFRS 部分处理结果展示

在本子小节，我们将分别展示 CLEAN 前后天线时域数据（图 6、图 11），频域数据（图 8，图 13），流量密度随频率分布的曲线（图 9），CLEAN 后 model 信号（图 12），互相关功率谱（图 10、图 14），相关系数随迭代次数变化曲线（图 15），并对相应的结果进行说明。

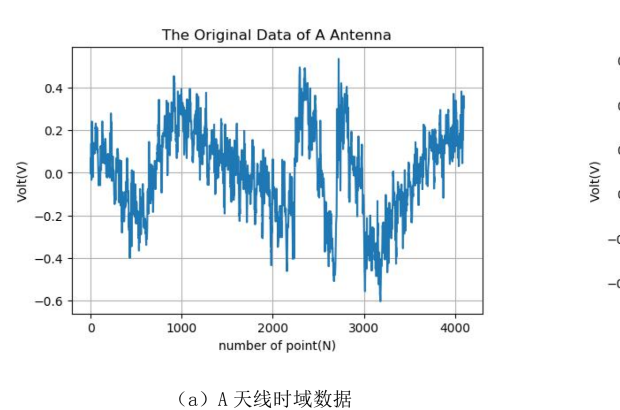


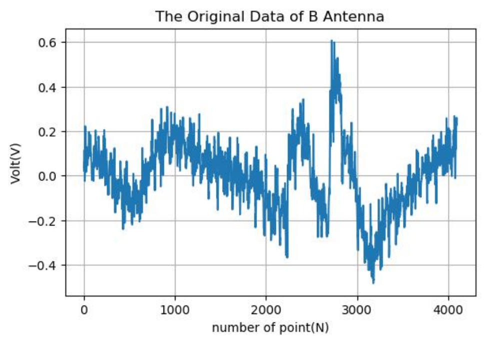
（a）A 天线时域数据　　（b）B 天线时域数据

（c）C 天线时域数据　　（d）D 天线时域数据

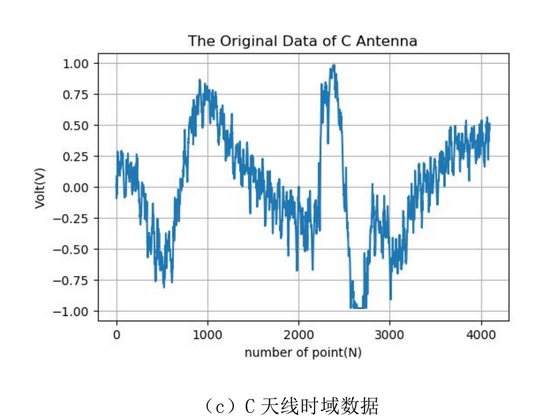
图6 A、B、C、D天线CLEAN前的第23月昼第一道原始时域数据

Fig.6 The first time-domain data, before CLEAN, of antenna A, B, C, and D on the 23rd moon day. (a)Time domain data of antenna A; (b) Time domain data of antenna B; (c) Time domain data of antenna C; (d) Time domain data of antenna D

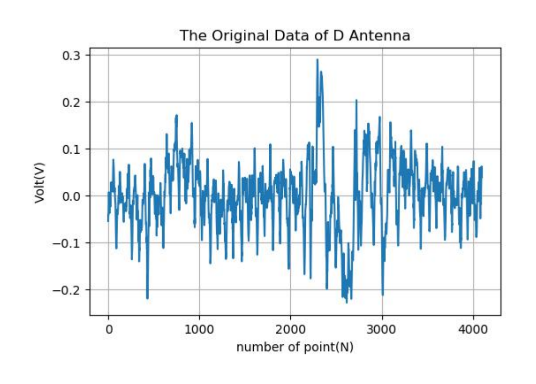




（a）A 天线时域数据　　（b）B 天线时域数据



（c）C 天线时域数据



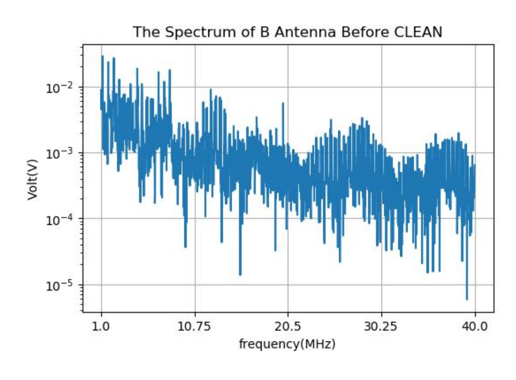
（d）D 天线时域数据

图7 A、B、C、D天线CLEAN前的第23月昼第二道原始时域数据

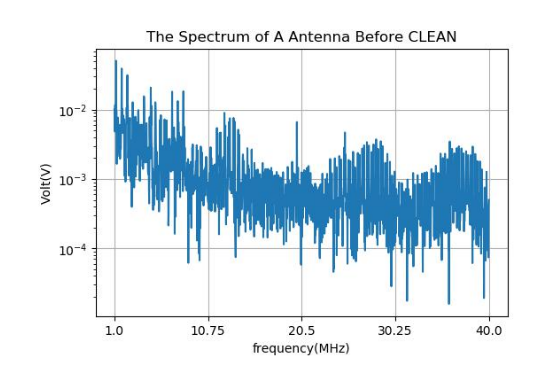
Fig.7 The second time domain data, before CLEAN, of antenna A, B, C, and D on the 23rd moon day. (a)Time domain data of antenna A; (b) Time domain data of antenna B; (c) Time domain data of antenna C; (d) Time domain data of antenna D

从 CLEAN 前的时域数据（图 6）中可以看到，A 天线和 B 天线的信号值范围基本接近，均为-0.5V～0.5V 左右。C 天线数据比 A、B 天线略大。三个天线的信号随时间变化的结构具有较高的相似性，A、B 之间的相关系数高达 0.965，B、C 之间的相关系数为 0.823，A、C 之间的相关系数为 0.915。同时，数据中存在的来回震荡结构主要成分不是接收机噪声，因为接收机噪声之间无相关性。A、B、C 天线信号的均值分别为-0.0149V、-0.0143V 和-0.0032V，均接近于 0。它们的标准差σ分别为 0.189V、0.138V 和 0.409V，存在明显的差别。这种差别可能是由于三天线的增益不同，或者平台干扰在三天线上的投射系数不一致，具体原因还有赖于后续细致的数据分析。从 CLEAN 前第二道时域数据（图 7）中可以看到，第一道数据与第二道数据时域信号有较为明显的差别。

从四根天线的数据来看，发现 D 天线的响应值比其余三根天线小很多，这是由于 D 天线长度只有 20cm，比 A、B、C 长度（5m）小很多造成的。同时，D 天线信号与其余三根天线信号在形状上也具有明显的差别，说明 D 天线的频谱响应明显异于其余三根天线，并且这种差异是非线性的。从天线之间的相关系数也可以看出，A、B、C 三根天线与 D 天线的数据相关系数分别为 0.440，0.397 和 0.298，显著低于 A、B、C 天线信号之间的相关系数。因此，我们放弃使用 D 天线数据来进行平台干扰信号的分离。



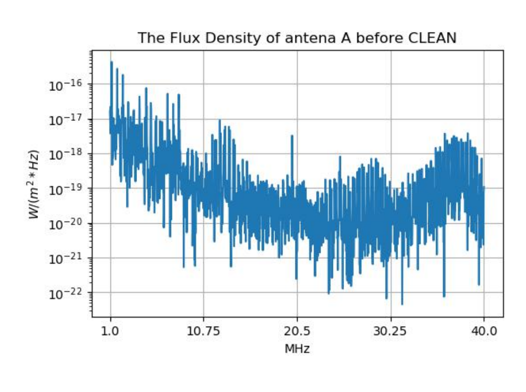
（a）A 天线时域数据 CLEAN 前的频谱



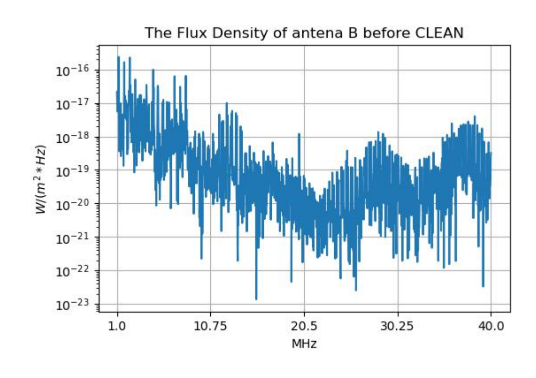
（b）B 天线时域数据 CLEAN 前的频谱

图8 A、B天线CLEAN前第23月昼第一道数据的频谱

Fig.8 The spectrum of the first data, before CLEAN, of antenna A and B on the 23rd moon day. (a) The spectrum, before CLEAN, of antenna A; (b) The spectrum, before CLEAN, of antenna B;



（a）A 天线 CLEAN 前流量密度



（b）B 天线 CLEAN 前流量密度

图9 A、B天线第23月昼第一道数据CLEAN前的流量密度

Fig.9 The first flux density data, before CLEAN, of antennas A and B on the 23rd moon day. (a)Flux density of antenna A before CLEAN;(b) Flux density of antenna B before CLEAN

从两天线 CLEAN 前的频谱（图 8）可以看到，频谱的幅度谱强度分布在$10^{-1}V$~$10^{-5}V$之间，低频成分强度较大，并且频谱中呈现出复杂的结构。将此时的电压频谱转化为流量密度频谱（图 9）后，发现两天线较低频段的流量密度峰值可超过$10^{-16}W/(m^2\cdot Hz)$，比太阳爆发信号的流量密度$10^{-18}W/(m^2\cdot Hz)$还要大 2 个量级。说明直接读取的时域信号如果不经过去除平台干扰的处理，我们很难发现宇宙射电爆发信号。

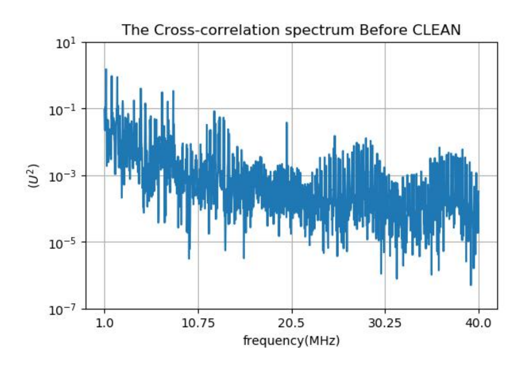
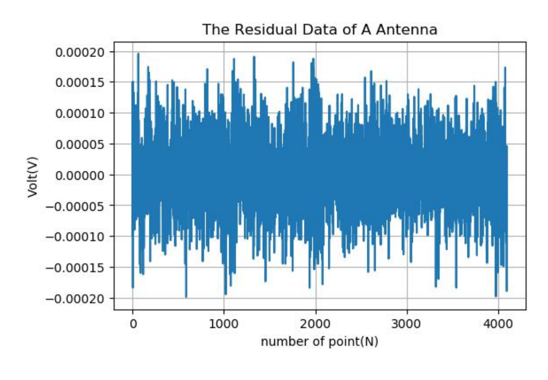


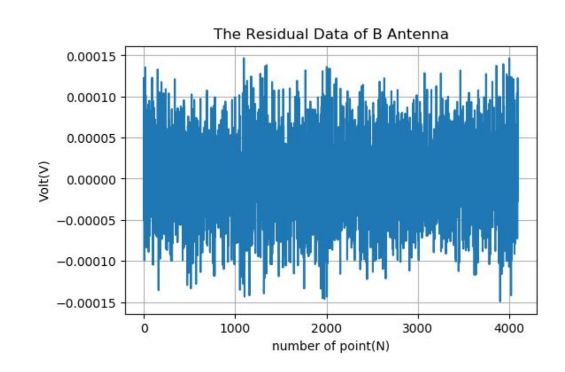
图 10 A、B 天线 CLEAN 前的互相关功率谱

Fig.10 Cross power spectrum of antennas A and B before CLEAN

两天线在 CLEAN 前的互相关功率谱（图 10）显示其中存在一些较为复杂的结构。此外，与单天线频谱相比较，两天线互相关功率谱的信噪比更高。对比图 10 和图 8，可以发现单天线频谱中，低频的峰值大约比高频的平均值高了约 1 个多数量级，而在两天线互相关功率谱中，低频的峰值大约比高频的平均值高了约 2 个多数量级。因此，从互相关功率谱出发，更有利于把平台干扰信号给提取出来。



（a）A 天线 CLEAN 后的残余信号

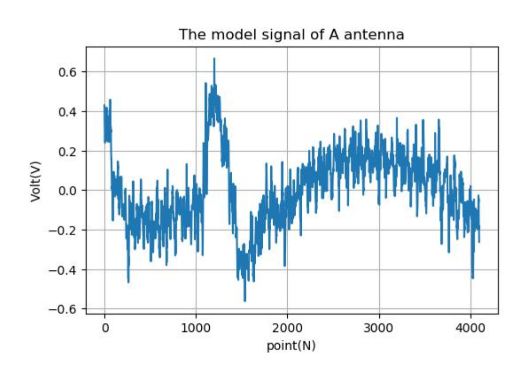


（b）B 天线 CLEAN 后的残余信号

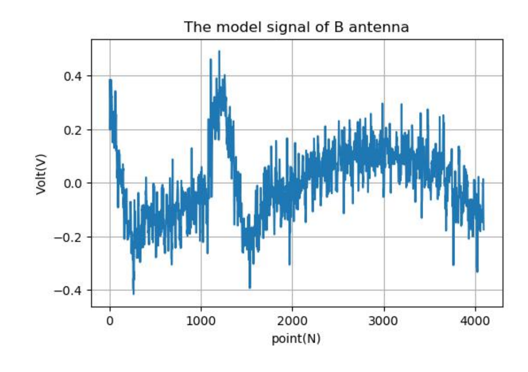
图11 A、B天线CLEAN后的残余信号

Fig.11 The after-CLEAN residual signal of antennas A and B.(a)The residual signal of antenna A after CLEAN; (b) The residual signal of antenna B after CLEAN;

我们进行 CLEAN 时选取的δ增益因子为 0.2，迭代次数设置为 10 万次。从经过 CLEAN 之后，天线 A 和天线 B 的残余信号显示在图 8 中。可以看出，CLEAN 前天线时域信号中存在的一些强峰结构已经消失了。CLEAN 之后天线 A 数据的σ为 $6.58\times10^{-5}V$，天线 B 数据的σ为 $5.36\times10^{-5}V$，可见天线数据的波动性已经得到极大地减小。天线信号现在在不同时间段上的幅度近似相等，这与接收机的白噪声本底的特点相符合。



（a）A 天线 CLEAN 后的模型信号

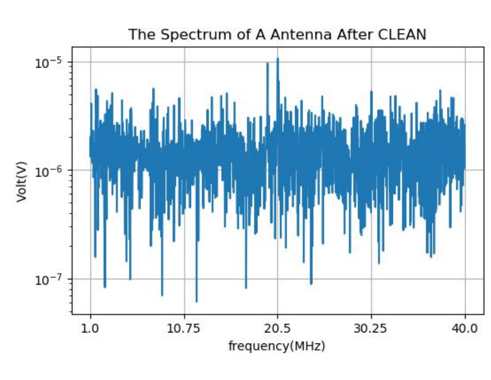


（b）B 天线 CLEAN 后的模型信号

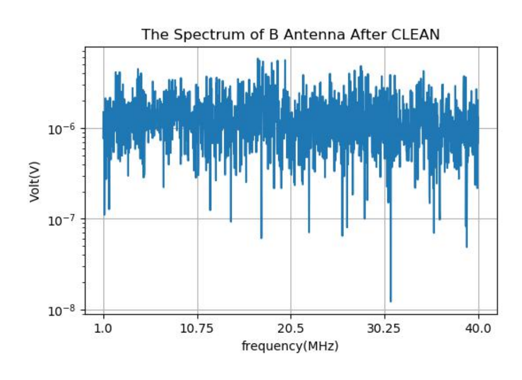
图12 A、B天线CLEAN后的模型信号

Fig.12 The model signal of antenna A and B after CLEAN. (a)Model signal of antenna A after CLEAN; (b) Model signal of antenna B after CLEAN;

从两天线 CLEAN 之后得到的模型信号分布（图 12）中可以看出，CLEAN 之后的 model 信号与初始信号很接近，说明原始的信号中大部分是平台干扰信号。同时，两天线 model 信号的相关系数为 0.966，接近于 1，也说明 model 信号中大部分是相关的平台干扰信号。



（a）A 天线 CLEAN 后频谱



（b）B 天线 CLEAN 后频谱

图13 A、B天线直接CLEAN后频谱

Fig.13 Spectrum of antennas A and B after CLEAN. (a)Spectrum of antenna A after CLEAN; (b) Spectrum of antenna B after CLEAN;

CLEAN 之后两天线残余信号的频谱（图 13）幅度峰值由原来的约$10^{-1}V$ 降低到约$10^{-5}V$，降低约 4 个数量级。残余信号的频谱也不再有明显的频谱结构，在各个频率分量的强度大致相等，可以看作白噪声谱。这说明，天线信号中的强平台干扰成分得到了非常有效的去除。

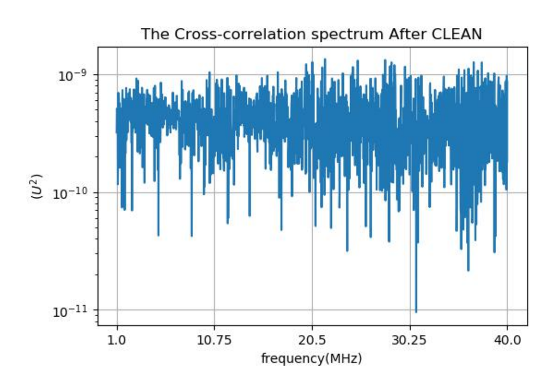


图 14 A、B 天线 CLEAN 后的末态互相关功率谱

Fig.14 The cross-power spectrum of antennas A and B after CLEAN

CLEAN 后的互相关功率谱（图 14）同样印证了算法的有效性。可以看出，两天线的互相关功率谱中已经没有明显的峰值结构，各个频段的互相关功率谱峰值都接近于$10^{-9}V^2$，对比图 7 中的结果，整整降低了约 9 个数量级。

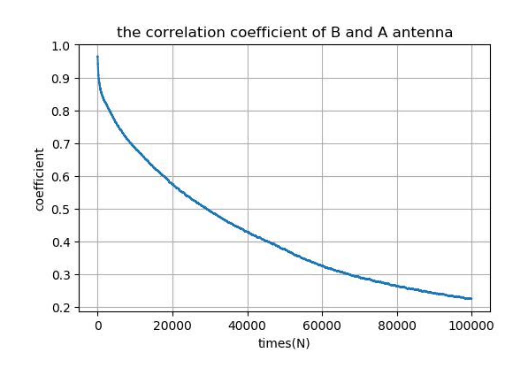


图15 A、B天线CLEAN时相关系数随迭代次数分布曲线

Fig.15 Distribution curve of correlation coefficient between antennas A and B vs. CLEAN iteration.

我们给出了两天线信号之间的相关系数随 CLEAN 迭代次数变化的曲线（图 15）。CLEAN 之前初始的相关系数为 0.965，由强平台干扰信号所主导；CLEAN 十万次之后的相关系数接近于 0.2，根据此相关系数计算出的信噪比，可以反映出此时残余信号中的平台干扰信号和接收机噪声基本在同一个水平上，此时信号由接收机噪声所主导。当然，CLEAN 结束之后，两天线信号的相关系数并没有显著接近于 0。这说明，残余信号中可能包含有来自于宇宙的低频射电信号，以及没有彻底清除的平台干扰信号。

4.3 CLEAN 方法效果分析

我们假设A、B两个天线的时域数据分别表示为$f^A(t)=S(t)+N^A(t)$和$f^B(t)=S(t)+N^B(t)$，其中$S(t)$为完全相同或相关的信号成分，设其均值为0，σ记为σ_S，$N^A(t)$和$N^B(t)$分别表示A和B天线的接收机噪声信号，其均值也都为0，σ都记为σ_N。但是，$N^A(t)$和$N^B(t)$是互相独立的。相关系数采用线性相关系数的定义$r=\frac{cov(A,B)}{\sigma_A\cdot\sigma_B}$，信噪比SNR定义为$\sigma_S/\sigma_N$。通过推导信噪比与相关系数之间的关系，可以得到如下公式：

$$r=\frac{SNR^2}{SNR^2+1} \quad (5)$$

当相关系数从最初的 0.965 降低到最后接近 0.2 时，由公式（5）计算得到的信噪比 SNR 从 5.3 降低到了 0.5。说明在末态残余信号中，相干的信号成分和不相干的噪声成分几乎在同一个水平上。

通过将末态信号转化为流量密度，得到 B 天线末态信号的流量密度随频率的分布（图 15）。从中可见，末态流量密度基本上降低到$10^{-23}\mathrm{W/(m^2\cdot Hz)}$以下，与太阳宁静时的信号水平相当。后续可以在此基础上对强射电爆发进行研究。

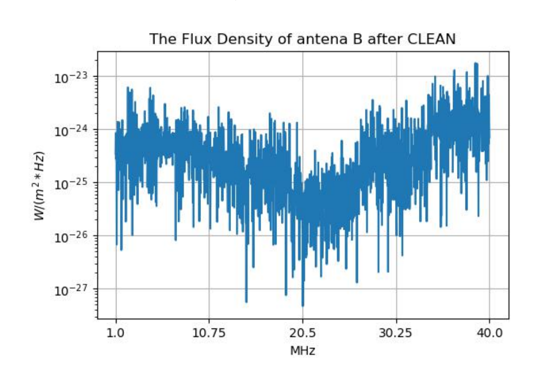


图 15 B 天线 CLEAN 后残余信号的流量密度

Fig.15 The residual flux density of antenna B after CLEAN

5 讨论

经过 CLEAN 算法处理之后，嫦娥四号低频射电频谱仪各天线的探测灵敏度（未积分）从$10^{-15}\mathrm{W/(m^2\cdot Hz)}$提高到$10^{-23}\mathrm{W/(m^2\cdot Hz)}$，整整 8 个数量级。在此灵敏度下，可以对太阳爆发进行监测。太阳爆发的流量密度一般在$10^{-18}\mathrm{W/(m^2\cdot Hz)}$，极端爆发可以达到$10^{-15}\mathrm{W/(m^2\cdot Hz)}$。除此之外，这样的灵敏度也足以探测部分强的快速射电爆(Fast Radio Burst, FRB)，其辐射流量密度可以达到几十万央斯基（Jy），$1\mathrm{Jy}=10^{-26}\mathrm{W/(m^2\cdot Hz)}$。如果再对时域数据进行积分平均，那么可以进一步提高探测灵敏度。例如，如果对一道 4096 个数据进行积分平均，则灵敏度可以额外提高 64 倍。这将大大拓展低频射电频谱仪的科学探测目标。

一系列用 CLEAN 算法分离出来的平台干扰信号也很有用。首先是被用来分析平台干扰信号的频谱、空间和时间变化特征，并确定其来源，以便在今后开发类似设备的时候，改进电磁兼容设计，大幅度降低平台干扰水平。其次，在全面了解平台干扰信号特征的基础上，充分利用这些知识，发现那些特征异常的信号，这些异常信号往往来自于宇宙。

分离并抑制干扰信号的最终目的是进行有效的天文观测。那么，如何在强烈的平台干扰背景下寻找并确认宇宙射电信号呢？主要的思路是寻找异常信号。例如，频谱异常。平台干扰往往有固定的一组频谱特征，一旦有例外的频谱信号出现，可能意味着记录到了来自宇宙的射电爆发信号。如果这些频谱还具备色散等特征，那么几乎可以肯定它来自宇宙。再例如，A、B、C 天线信号的强度比例系数。若平台干扰信号产生的空间分布是固定的，那么不管其在时域如何变化，信号之间的强度比例系数也是固定的。倘若这些比例系数有明显的改变，则往往意味着天线接收到了来自宇宙的强射电爆发信号。若射电爆发信号的时标长达几天（此处 1 天=24 小时）的量级，那么这些比例系数还会随着月球自转，发生规律性变化，有助于更加可靠的确定爆发信号的来源。当然，具体如何搜寻异常信号，还得依赖于后续细致的数据分析工作。

本论文提出的方法也可以应用到其它射电观测设备，尤其是地基的综合射电干涉阵。目前地基干涉阵受到越来越多的来自地面设备以及空间卫星的无线电干扰。借助 CLEAN 算法，我们利用俩俩天线之间的互相关功率谱，在信号层面把干扰信号提取出来，再对残余数据做相关获得可见度数据，对可见度数据作成图等处理工作。同样，也可以对提取出的干扰信号进行作类似的操作，获得其详细的频谱、空间分布、时域变化等信息。

6 结论

本论文提出一种方法，它从两组信号的相关性出发，基于 CLEAN 算法，并借助互相关功率谱、傅立叶级数展开等工具，把低频射电频谱仪 A、B、C 天线的时域观测数据分为强相关

的 CLEAN 模型信号和部分相关的残余信号两个部分。其中 CLEAN 模型信号主要由平台干扰信号和可能的低频强射电爆发组成；残余信号由接收机噪声、未扣除干净的平台干扰信号和常规的低频射电信号组成。将该方法应用到实际数据中，结果表明嫦娥四号低频射电频谱仪的未积分灵敏度可以提高约 8 个数量级，达到 10^{-23}W/(m^2*Hz)水平。

在此基础上，基于对平台干扰信号中确定性成分和宽带随机成分的分别处理，再借助低频射电爆发信号和平台干扰信号在功率谱上不同的表现，以及常规低频射电信号受月球自转调制等信息，将来工作的重点是进一步分析 CLEAN 模型信号和残余信号，以期搜索低频强射电爆发信号，乃至对全天区进行粗略的成像。

Using CLEAN Algorithm to Suppress the Interference of Chang'e-4 Low-frequency Radio Spectrum Analyzer

Liu Chendi[1,2,3],Zhou Jianfeng[4],Su Yan[1,2,3]

(1. National Astronomical Observatories, China Academy of Sciences, Beijing 100101, China;2. University of China Academy of Sciences, Beijing 100049, China;3.Key Laboratory of Lunar and Deep-Space Exploration, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China;4. Department of Engineering Physics, Tsinghua University, Beijing 100084, China)

Abstract: The Chang'e-4 low-frequency radio spectrometer is placed on the back of the moon, and its natural observation conditions are unique. However, the Chang'e-4 platform has strong interference of about $10^{-15}W/(m^2 \cdot Hz)$ level, and there are obvious differences in the interference in each channel of time domain data, which greatly weakens the astronomical observation sensitivity of the low-frequency radio spectrum analyzer. To this end, starting from the correlation of the two sets of signals, this paper proposes to divide the time-domain observation data of the low-frequency radio spectrum analyzer A, B, and C antennas into strongly correlated CLEAN model signals and partial correlated residual signal based on the CLEAN algorithm, with the help of cross-correlation power spectrum, Fourier series expansion and other tools. The CLEAN model signal is mainly composed of platform interference signals and possible low-frequency strong radio bursts; the residual signal is composed of receiver noise, undeducted platform interference signals and conventional low-frequency radio signals. Applying this method to actual data, the results show that the unintegrated sensitivity of the Chang'e-4 low-frequency radio spectrum analyzer can be increased by about 8 orders of magnitude, reaching a level of $10^{-23}W/(m^2 \cdot Hz)$. On this basis, based on the separate processing of the deterministic components and broadband random components in the platform interference signal, and then relying on the different performance of the low frequency radio burst signal and the platform interference signal in the power spectrum, and the conventional low frequency radio astronomy signal modulated by the moon rotation And other information, the focus of future scientific analysis work is to further process the CLEAN model signal and residual signal, with a view to discovering low-frequency strong radio astronomy burst signals, and even roughly imaging the entire sky.

Key words: Chang'e-4 low-frequency radio spectrum analyzer; Platform interference removal; CLEAN algorithm

参考文献

[1] 梅丽, 苏彦, 周建锋. 极低频射电天文观测现状与未来发展[J]. 天文研究与技术, 2018, 15(2):127-149.

[2] Leshem A , Veen A V D . Detection and blanking of GSM interference in radio-astronomical observations[C]// IEEE Workshop on Signal Processing Advances in Wireless Communications. IEEE, 2002.
[3] Fridman P A , Baan W A . RFI mitigation methods in radio astronomy[J]. Journal of Astronomy & Astrophysics, 2001, 378(1):327-344.
[4] Ellingson S W , Bunton J D , Bell J F . Removal of the GLONASS C/A Signal from OH Spectral Line Observations Using a Parametric Modeling Technique[J]. The Astrophysical Journal Supplement Series, 2000, 135(1).
[5] Ellingson S W . Beamforming and interference canceling with very large wideband arrays[J]. IEEE Transactions on Antennas and Propagation, 2003, 51(6):1338-1346.
[6] Cecilia B , Bradley R F . A New Approach to Interference Excision in Radio Astronomy: Real-Time Adaptive Cancellation[J]. Astronomical Journal, 1998, 116(5):2598.
[7] Briggs F H , Bell J F , Kesteven M J . Removing radio interference from contaminated astronomical spectra using an independent reference signal and closure relations[J]. Astronomical Journal, 2000, 120(6):3351-3361.
[8] Leshem A , van der Veen, Alle - Jan, Boonstra, Albert - Jan. Multichannel interference mitigation methods in radio astronomy[J]. Astrophysical Journal Supplement, 2000, 131(1):355.
[9] Friedman, P. [IEEE Comput. Soc. Press 8th Workshop on Statistical Signal and Array Processing - Corfu, Greece (24-26 June 1996)] Proceedings of 8th Workshop on Statistical Signal and Array Processing - A change point detection method for elimination of industrial interference in radio astronomy receivers[J]. 1996:264-266.
[10] Weber R , Faye C , Biraud F , et al. Spectral detector for interference time blanking using quantized correlator[J]. Astronomy & Astrophysics Supplement, 1997, 126(1):161-167.
[11] Maslakovic S , Linscott I R , Oslick M , et al. [IEEE Third International Symposium on Time-Frequency and Time-Scale Analysis (TFTS-96) - Paris, France (18-21 June 1996)] Proceedings of Third International Symposium on Time-Frequency and Time-Scale Analysis (TFTS-96) - Excising radio frequency interference using the discrete wavelet transform[J]. 1996:349-352.
[12] Kasper B L , Chute F S , Routledge D . Excising terrestrial radio interference in low frequency radio astronomy[J]. Monthly Notices of the Royal Astronomical Society(2):345-354.
[13] Cecilia B , Bradley R F . A New Approach to Interference Excision in Radio Astronomy: Real-Time Adaptive Cancellation[J]. Astronomical Journal, 1998, 116(5):2598.
[14] Baan W A , Fridman P A , Millenaar R P . Radio Frequency Interference Mitigation at the Westerbork Synthesis Radio Telescope: Algorithms, Test Observations, and System Implementation[J]. Astronomical Journal, 2004, 128(2):933-949.
[15] Offringa A R , De B A G , Biehl M , et al. Post-correlation radio frequency interference classification methods[J]. Monthly Notices of the Royal Astronomical Society, 2010, 405(1):155-167.

[16] Zhao, J, Zou, et al. WindSat Radio-Frequency Interference Signature and Its Identification Over Greenland and Antarctic[J]. IEEE Transactions on Geoscience & Remote Sensing, 2013, 51(9):4830-4839.
[17] C. J. Wolfaardt, Machine learning approach to radio frequency interference (rfi) classification in radio astronomy[D]. Stellenbosch University (2016).
[18] 张韬, 苏彦. 嫦娥四号低频射电频谱仪降低背景噪声方法的研究[J]. 天文研究与技术, 2019, 16(3):312-320.
[19] Bastian, T. S ,Benz, et al. Radio emission from solar flares.[J]. Annual Review of Astronomy & Astrophysics, 1998, 36:131-188.
[20] Bochenek C D, Ravi V, Belov K V, et al. A fast radio burst associated with a Galactic magnetar[J]. Nature, 2020, 587(7832): 59-62.
[21] Hogbom J A. Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines[J]. Astronony Astrophys Suppl, 1974, 15:417.
[22] Heslop D, Dekkers M J. Spectral analysis of unevenly spaced climatic time series using CLEAN: signal recovery and derivation of significance levels using a Monte Carlo simulation[J]. Physics of the earth and Planetary Interiors, 2002, 130(1-2): 103-116.
[23] Cousson R, Leclere Q, Pallas M A, et al. A time domain CLEAN approach for the identification of acoustic moving sources[J]. Journal of Sound and Vibration, 2019, 443: 47-62.
[24] Kulpa K . The CLEAN type algorithms for radar signal processing[M]. IEEE, 2008.
[25] 朱新颖，探月工程嫦娥四号任务着陆器低频射电频谱仪数据预处理方法设计[R].北京.2018.